THE RAPID SELECTOR

by RALPH R. SHAW

Librarian, United States Department of Agriculture

Introduction

THE Rapid Selector is the culmination of many years of development and experimentation, in part directly connected with it, in part in the fields of organization of knowledge, electronics, mechanics, and other fields.

The first practical application of electronics to selection of data on film appears to have been made some twenty years ago by Dr. E. Goldberg, who at the time was with Zeiss in Dresden. His invention was protected by United States Patent No. 1,838,389, issued on 29 December 1931.

The basic principles of organization of knowledge applied in the present machine were developed by Dr. Vannevar Bush at the Massachusetts Institute of Technology more than ten years ago, and, while many changes were made in operating details in designing the Rapid Selector, its basic electronic system should be credited to Dr. Bush.

The funds for development of the Rapid Selector were provided by the Office of Technical Services of the Department of Commerce from an appropriation intended for development of new tools to advance the technological development of the nation and to result in wide public benefit. Mr. John C. Green, Director of the Office of Technical Services, not only supervised the contractual details, but also was of constant assistance in technical aspects of the development of the work.

The Engineering Research Associates of St. Paul, Minnesota, selected and designed the electrical, optical, and mechanical components, and assembled this first machine.

The author, with Dr. Bush's permission to use his concept for the public good, developed the programme, devised the method of mounting the coding system so that any size document could be taken without changing the size of the dot codes, devised the triggering dot scheme so that multiple codes could be used in connexion with a single reproduction of the text data, and supervised the design, construction, and testing of the machine.

Principles of electronic selection

Electronic selectors are merely devices which use electronic controls to achieve purposes which can only be achieved more slowly by known means or which are of such magnitude that they cannot be achieved at all by any other known means. In its essence, electrons travelling at the speed of light are used to throw switches at high speed. Mechanical switches, depending on human factors and physical effort, are relatively slow; automatic switches, which involve movement of physical objects or the making of physical

© Emerald Backfiles 2007

contacts, have considerable time-lags. Electronic switches, while having some time-lag, can effect the closing of a circuit in a thousandth or a hundred-thousandth part of the time required to achieve the same result by manual or mechanical means.

The Rapid Selector, taking advantage of this principle of high-speed electronic activation, makes possible the handling of data and the reproduction of results in tiny fractions of the time required by other known means.

The fundamental principle of operation of the Rapid Selector is no different from the operation of the familiar 'electric eye' door-opening device, in which a beam of light shines on a photocell. A photocell is a unit which is capable of converting the energy received in the form of light into an electrical current. As the current emitted by a photocell is tiny and must be amplified before it can serve a useful purpose, so a photocell, for all practical purposes, must be associated with an amplifier to strengthen the current received from the photocell. This amplified current, in the case of a door-opening device, serves in effect as a brake, i.e. so long as the current flows the door is kept closed. However, when someone walks through the light beam he interrupts the light beam for a short period. The instant the light beam is interrupted the photocell stops putting out its electric current because it no longer receives energy in the form of light to convert into electrical current. Thus the current which is acting as a brake cuts off and the door-opening mechanism is permitted to open the door.

The same principle is used in the electronic selector, except that the pattern of light is coded to identify the person or subject for whom the door is to be opened, so that the current from the photocell cuts off only for a preselected code. When the current from the photocell is interrupted, the electronic brake goes off. This permits a flash lamp to spark for two-millionths of a second, producing, through an auxiliary camera, an enlarged print of the data on the microfilm which corresponds to the code which has been selected.

Objectives of electronic selection

There appear to be at least three broad areas in which electronic rapid selection and reproduction may be useful.

The first of these quite obviously appears to be the use of the Rapid Selector as a means for reducing storage space while permitting rapid finding and reproduction. In the first model of the Rapid Selector 72,000 frames of material are stored on a 2,000-foot reel of film. These frames may be of any size, up to a legal-size sheet. Seventy-two thousand sheets of paper would occupy approximately 300 cubic feet of space in a normal filing room, 8 feet high. In addition to the text pages stored on the film, however, the other half of the film provides space for as many as 430,000 index entries. Four hundred and thirty thousand index entries stored normally on 3 in. × 5 in. index cards would occupy approximately 800 cubic feet of space. Thus the text and accompanying index entries on a single roll of film would require approximately 1,100 cubic feet of space, as against approximately a quarter

© Emerald Backfiles 2007

of a cubic foot required for the film form. This means that the Rapid Selector uses approximately two-hundredths of 1 per cent. of the space required for comparable storage and indexing of the originals. Furthermore, the selector provides for instantaneous reproduction of copies, and that is not provided in conventional storage.

A section of wall shelving of standard height and about 12 feet long could hold approximately 20 million pages of text and approximately 120 million index entries, as filmed for the first model of the Rapid Selector. There is every reason to believe, on the basis of the tests made to date, that the storage capacity can be increased by at least 25 per cent. without changing any of the characteristics of the selector or any of the electronic circuits, and with only modest changes in the recording mechanism.

Use of microfilm for storage of large amounts of material in a minimum space is not a new idea. However, before the development of the Rapid Selector, microfilm could not be used economically for storage of live records. It was confined to more or less dead storage of long runs of little-used materials. Even for that type of use the material had to be very carefully prearranged to permit finding. The use of microfilm for live materials, into which new frames must be inserted from time to time, was quite impracticable, and its use for random storage of information was utterly impossible because of the difficulty of locating material on the film.

However, with the Rapid Selector as it exists or as it may be developed, it appears that it should be quite feasible to film letters or other data at random as received, coding the letters at least as thoroughly as they would be coded for normal filing, so that in some cases the selector may replace files and indexes to files. In the prototype, searching of material takes place at the rate of 78,000 subject codes per minute. Improvement of this searching speed to at least 120,000 subjects per minute is now in sight.

A second field of application of rapid selection and reproduction would appear to be the field of cumulative indexing of materials from various sources, rather than control of material in a single file. A thorough search of the literature of chemistry, for example, may require the handling of dozens, or even hundreds, of volumes of indexes. Once this material from all the indexes is coded into film, the time required for electronic searching should be a tiny fraction of the time required for manual searching.

If *Chemical Abstracts* were available in coded form in the selector, hundreds if not thousands of hours spent by chemists in the mechanical job of searching would be saved. At the present time a chemist may have to spend anything from half a day to several weeks searching *Chemical Abstracts* before he can undertake a new research project. When *Chemical Abstracts* has been coded into the film, the chemist would still have to do the intellectual job of determining what phases or fields should be searched. He would then merely consult a code book comparable to the *Decennial Index of Chemical Abstracts* and send the numbers he wants searched to the machine operator. The clerk operating the machine would send him copies of all the abstracts under these subjects in a very few minutes, and the mechanical part of the

© Emerald Backfiles 2007

searching job would be eliminated, thus freeing professional time for research.

The third area of possible usefulness of the machine appears to be the possibility of improvement in the quality of organization of knowledge both for administrative routines and for communication among scientists.

The greatest limiting factors in the organization of knowledge appear to have been purely physical factors of cost and bulk. In fact, it might even be said that substantially all of our schemes for organizing knowledge have been based to a certain extent upon physical factors or at least have been affected by compromises based upon physical factors.

Take *Chemical Abstracts* as an example. The cost of preparing the indexes is a very large part of the cost of producing *Chemical Abstracts.* Almost anyone who has searched *Chemical Abstracts* will agree that the indexing is not as detailed as might be desirable, and the editorial staff of *Chemical Abstracts* would be among the first to agree that they cannot physically produce an index to every idea in every abstract. If that were attempted, the index to the abstracts would be several times greater than the volumes of abstracts, and the cost would be prohibitive. Furthermore, it is doubtful if such detailed indexing would improve usability so long as the user must handle the volumes to determine whether the concept indexed actually applies to his problem. It might very well be that under present methods for handling materials a more detailed index to *Chemical Abstracts* would result in more rather than less time in searching because of the physical part of the searching required. Furthermore, the cost of such an index would appear to make it utterly impossible of achievement.

The limitations on intensity of indexing, working in both directions, i.e. with the cost factor tending to keep the number of index entries down and the effort to indicate all possible relationships tending to increase the number of entries per abstract, must result in the best compromise which can be made under a given set of circumstances. (And it should be noted that the compromise now effected in *Chemical Abstracts* has probably resulted in the best indexing which has been achieved anywhere to date.)

Similarly, in such collections as the Patent Office search files, an attempt is made to place a copy of each patent physically in each group in which it may be of interest. However, with several million patents, if each is placed in only 8 or 10 places, a file of almost unworkable proportions results. Since there may be 10 or 20 different ideas of interest in each patent, if the indexing of knowledge were carried to its ultimate conclusion, a complete searching file of U.S. patents alone might include several hundreds of millions of copies of patents. Again, the result is a compromise which leaves a great deal to the imagination, the energy, and the time available to the searcher.

A similar physical problem exists in the organization of books in libraries, or of technical documents in data files, for that matter. The grouping of books on chemistry under a decimal number does not bring together everything in the collection on the subject of chemistry. Everyone would agree that a very large amount of valuable data on any aspect of chemistry will be

© Emerald Backfiles 2007

found scattered in journals. It is not physically possible to place a copy of every article, or of every page of every article, under every subject under which it may be sought. Thus a very rough compromise is adopted which places a small portion of the material on any subject in one place on a fairly rough basis. This is necessary to avoid running to dozens of places every time one wants to consult the general text-books on chemistry. It would be impossible to run to a hundred places every time anything on the subject is desired. Thus, classification of books in libraries, or the data in technical data files, would appear to have been little more, at best, than a compromise based primarily on physical considerations, rather than an attempt to organize the intellectual content of the physical objects. This explains the need for indexes and abstract journals which carry the organization of the intellectual content one step farther by supplying index entries instead of complete reports to stand in the physical locations. Physical considerations, however, of even these representations have limited the availability of the intellectual content of the materials.

The Rapid Selector, through its compression of the physical data and its ability to run to hundreds of thousands of places in less time than we can walk to one, takes us a large step farther in our efforts to deal with the intellectual content of literature by greatly reducing the importance of the physical factor. While even with electronic selection it is possible to postulate bulks of material so great that appreciable time and cost are involved, the physical limitations may be reduced so greatly that for a long time to come they can be considered negligible. Thus it should be possible to record every concept, no matter where it occurs, into the human record permanently, in such form that it may be produced more rapidly and at lower cost than we can now achieve for our limited coverage. Since the machine can run to hundreds of thousands of places faster than we can go to one or two, assuming proper development of the intellectual content by indexing to take advantage of the new range offered by the machine, searches which cannot now be undertaken should become routine, and the quality of searching become, theoretically at least, much greater.

In this respect the capabilities of the machine appear to be ahead of our thinking. This third area of application is stressed because it appears that, while useful results may be achieved merely by using the machine to do more speedily and more efficiently what we can now do, as indicated in the first and second areas of application discussed above, a really important contribution to the advancement of science will result only if we can rethink the methods of organization of knowledge to take full advantage of the new technique. In this respect it would appear that we need first to do some fundamental thinking and some operational research to determine what is really needed for the advancement of scientific communication and for the advancement of science. Until we know that, it appears doubtful whether we shall make fullest possible use of any mechanism.

The selector has been termed a 'thinking' machine or 'electronic brain'. Without more knowledge of what 'thinking' consists of, it is difficult to

© Emerald Backfiles 2007

say whether the selector thinks or does not think. Certainly in the common sense of the term the selector is not a thinking machine. It merely stores vast amounts of data and sorts and reproduces them in accordance with instructions given to the machine both in the coding and in the selection. All the machine ever does is to match black dots (or if you prefer, light dots) with complementary dots in an interrogating card. The only answer the machine ever gives is 'yes' or 'no'. If the answer is 'no', i.e. the dots do not match, the machine runs along and nothing happens. When the answer is 'yes', a flash lamp is triggered to make a copy of a text frame which is predetermined in the coding and which the machine cannot change. The only thing the machine does is to store and reproduce data in accordance with whatever instructions are given by human minds. The coding system can be based on a numerical classification scheme, alphabetical classification scheme, or no scheme at all. It is utterly immaterial in so far as the operation of the selector is concerned, and that, indeed, may be one of its greatest advantages, since it can accept any or all schemes of organization of knowledge and can, theoretically at least, use them all at once or separately, depending only upon the thinking that has gone into the design of the coding pattern and the interrogating card.

The ability of the machine to handle material organized according to any classification scheme or subject-heading scheme is of particular importance. Progress in the organization of vast amounts of technical data has been delayed because of the need for agreement upon a scheme of classification or listing of the material in conventional systems of indexing. Yet in most fields it has appeared almost impossible to achieve agreement upon the scheme of organization of knowledge which should be adopted universally. This is probably inevitable. Since the grouping of his own science in his own mind is probably an essential component of the genius of research, i.e. it is the way a scholar discovers new relationships, it would appear undesirable for all scientists to attempt to approach their specialities from the identical point of view, seeing everything in identical relationships.

In addition, the creation of a universal classification scheme would appear to require pre-vision of the future developments of the field, particularly if reclassification in a growing science is to be obviated. If we knew enough about the future development in any field to provide for such development there would be no further need for research. The use of classification schemes in the ordinary methods makes change in large files almost impossible because of the mechanical job of extracting cards and reclassifying or regrouping them. However, by the use of random numbering in the Rapid Selector, code numbers can be applied to the known concepts as organized in any given field by the specialist in the field, and as new concepts develop new code numbers can be assigned. Since these are not in any logical sequence, there would be no crowding of the subjects in the rubric and no need to leave room for growth within the various sequences of the classification scheme. In each field in which there is a subject-heading list or classification scheme currently acceptable to that field, all that need be done

© Emerald Backfiles 2007

is to number all the concepts and/or sub-concepts with a numbering stamp, making sure that synonyms receive the same number and that the same number is not used for two concepts. Then, when a new concept is added, regardless of where it fits in the classification scheme or subject-heading list, it gets the next consecutive number on the numbering stamp. Also, when new terminology develops, no change in headings is required; all that need be done is to add the new terminology to the code book, assigning the appropriate numbers previously used for that term under a different heading. This means that, regardless of which heading is used in approaching the code book, the number representing that subject could be located without cross references and without having to extract cards and change headings. These code numbers serve, in effect, as an index to the classification scheme or subject-heading scheme used by scholars in each discipline.

Furthermore, since multiple reels may be used for different sources or subject fields, the same numbers may be assigned to different concepts, provided only that the code book for each reel is kept separately from the code books from other reels, and the proper code book is used in connexion with the reel. Thus, for example, if it should prove too much trouble to sort out the synonyms in *Chemisches Zentralblatt* and match them with the English equivalents in *Chemical Abstracts*, this is easily solved by having *Chemisches Zentralblatt* on reels separate from those for *Chemical Abstracts* and making up a code book for *Chemisches Zentralblatt*, starting the numbers with one, just as is done in the case of *Chemical Abstracts*.

Under this scheme, the quality of indexing of the intellectual content would be no different from what it is now, except that such time and expense as would be saved from printing cumulative indexes in their present forms could be spent on more detailed indexing instead, so that eventually the intellectual work put into the original indexing might be of higher level for the same cost. The random assignment of numbers to subject headings can be done just as well with a classification scheme, such as the Brussels, by expressing the concepts as random numbers merely by stamping the number alongside the original concept.

Another field which might well be covered by the development of the electronic selector is the preparation of specialized indexing and abstracting journals automatically from the general indexing and abstracting journals. Much of the duplication in indexing at the present time is desirable duplication. The embryologist does not want to have to handle a general abstracting journal and to search through ten or fifteen different classifications to bring together the work on embryology. He would like to have all the abstracts on embryology in one place. Proper coding of the original abstract should make it possible to turn out special abstracting journals as is done in *Biological Abstracts* for example, merely by making a run of the whole issue for a month with the selector set to reproduce all abstracts which apply to that field of research, regardless of what other fields they fit. This should make it possible, then, to provide specialized abstracting services at relatively low cost without duplicating the original intellectual work, provided only

© Emerald Backfiles 2007

that the interests of all the groups concerned are taken into consideration when the original abstract is written. If this is done, then a single abstract in each field might well satisfy each specialist in his own field, as well as giving him a general abstracting journal to bring together all related material which he may wish to consult from time to time.

Summary of the objectives of electronic selection

It appears that there are three broad areas in which electronic selection and reproduction may be useful to administration and to scholarship. These are:

1. High-speed finding and reproduction of any type of file material stored in a minimum of space.
2. Cumulative indexing of technical data or other materials so that complete searches may be made of all known sources as required.
3. Organization of knowledge on a much higher level, on the basis of the intellectual content of material and the needs of research.

Note to the above

A definite report on the Rapid Selector has still to be written, but, for the sake of readers who are interested in the technical aspects of the machine, the following notes may be useful:

Office of Technical Services, U.S. Department of Commerce, Washington, D.C., publish a report describing the Rapid Selector in detail and accompanied by illustrations. This is PB 97535, price $2.50 per copy.

'New machine will scan, select and copy research data at high speed', *Library Journal*. 15 May 1948 (vol. 73, no. 10), pp. 797, 806. [Some notes by Mr. Ralph Shaw.]

'Optical punch card. Electronic brain searches literature with combination of microfilm and punch-card technique', *Chemical Industries*. August 1949 (vol. 65, no. 2), pp. 189–90. [A short description.]

'Photoelectric librarian', *Electronics*. September 1949 (vol. 22, no. 9), pp. 122, 158, 160, 162, 164, 166. [A fairly detailed description with illustrations.]

E. M. R. D.

© Emerald Backfiles 2007